**MapReduce on Yelp Dataset**

**How to run the code:**

Copy the Assignment2.jar into HDFS and use the following commands to run:

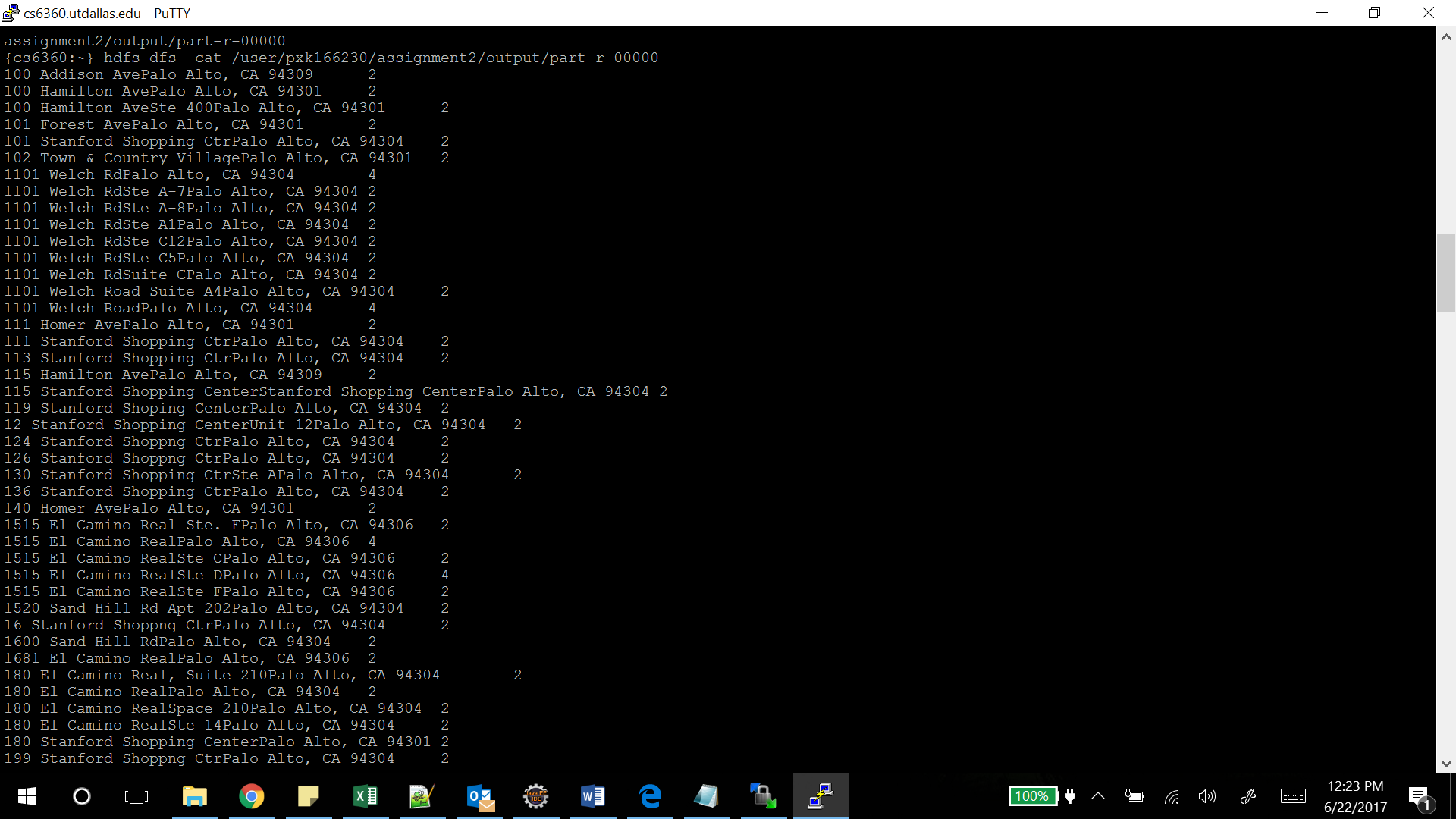
Q1:

*hadoop jar Assignment2.jar Q1 <input-path> <output-path>*

Example,

hadoop jar Assignment2.jar Q1 /yelp/business/business.csv hdfs://cshadoop1/user/pxk166230/assignment2/Q1/output

Screenshot of the output:



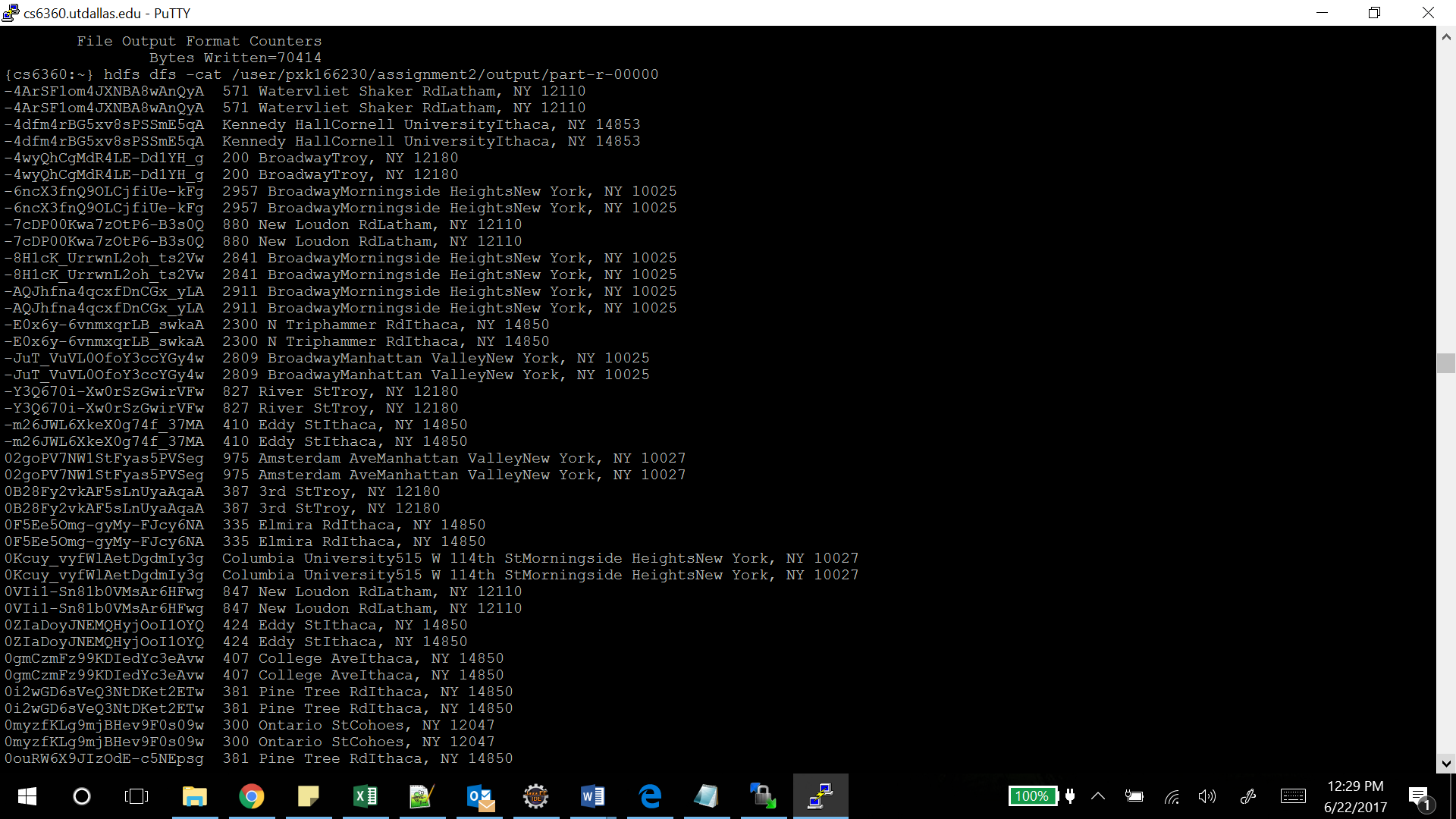
Q2:

*hadoop jar Assignment2.jar Q2 <input-path> <output-path>*

Example,

hadoop jar Assignment2.jar Q2 /yelp/business/business.csv hdfs://cshadoop1/user/pxk166230/assignment2/Q2/output

Screenshot of the output:



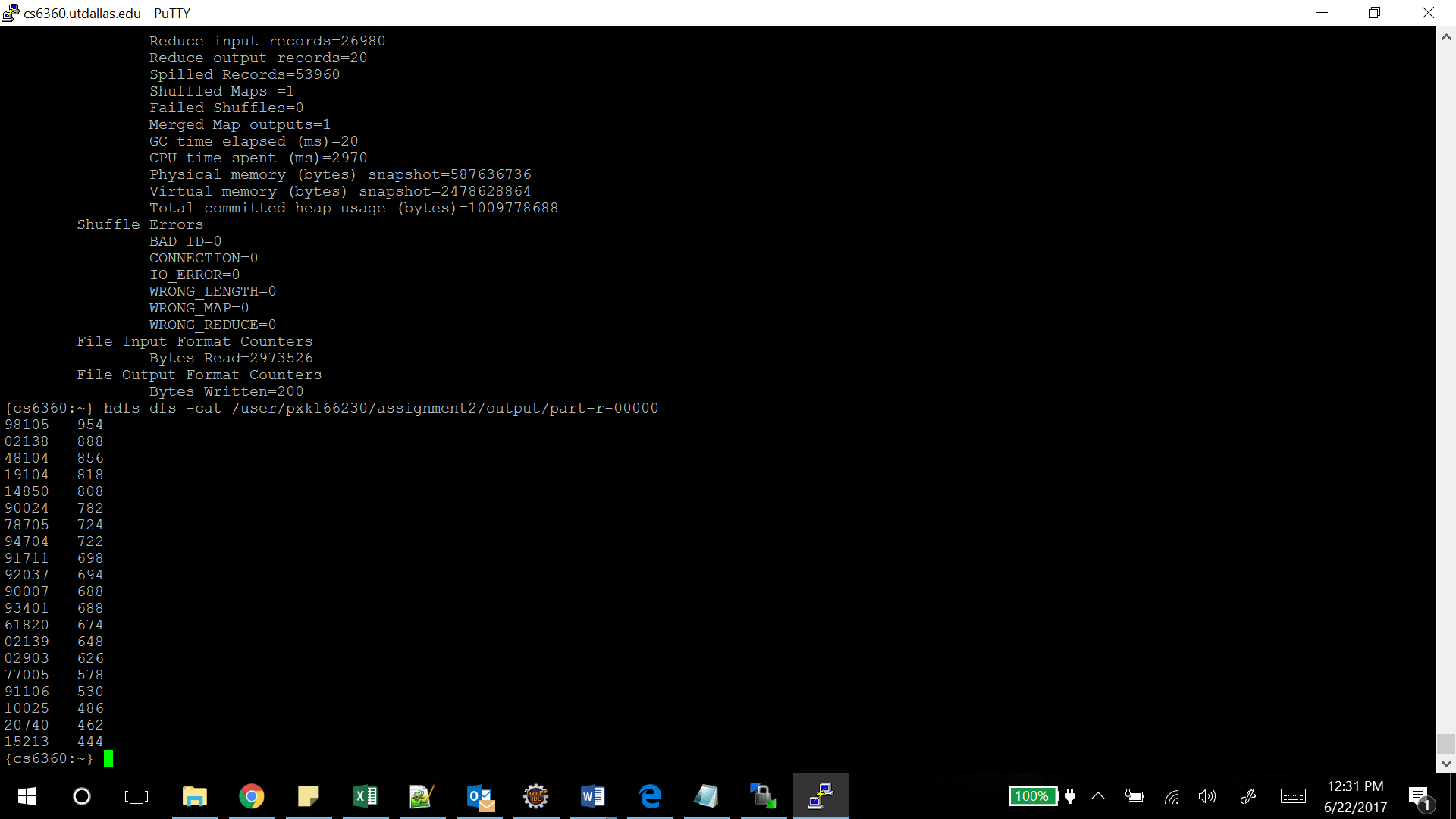
Q3:

*hadoop jar Assignment2.jar Q3 <input-path> <output-path>*

Example,

hadoop jar Assignment2.jar Q3 /yelp/business/business.csv hdfs://cshadoop1/user/pxk166230/assignment2/Q3/output

Screenshot of the output:



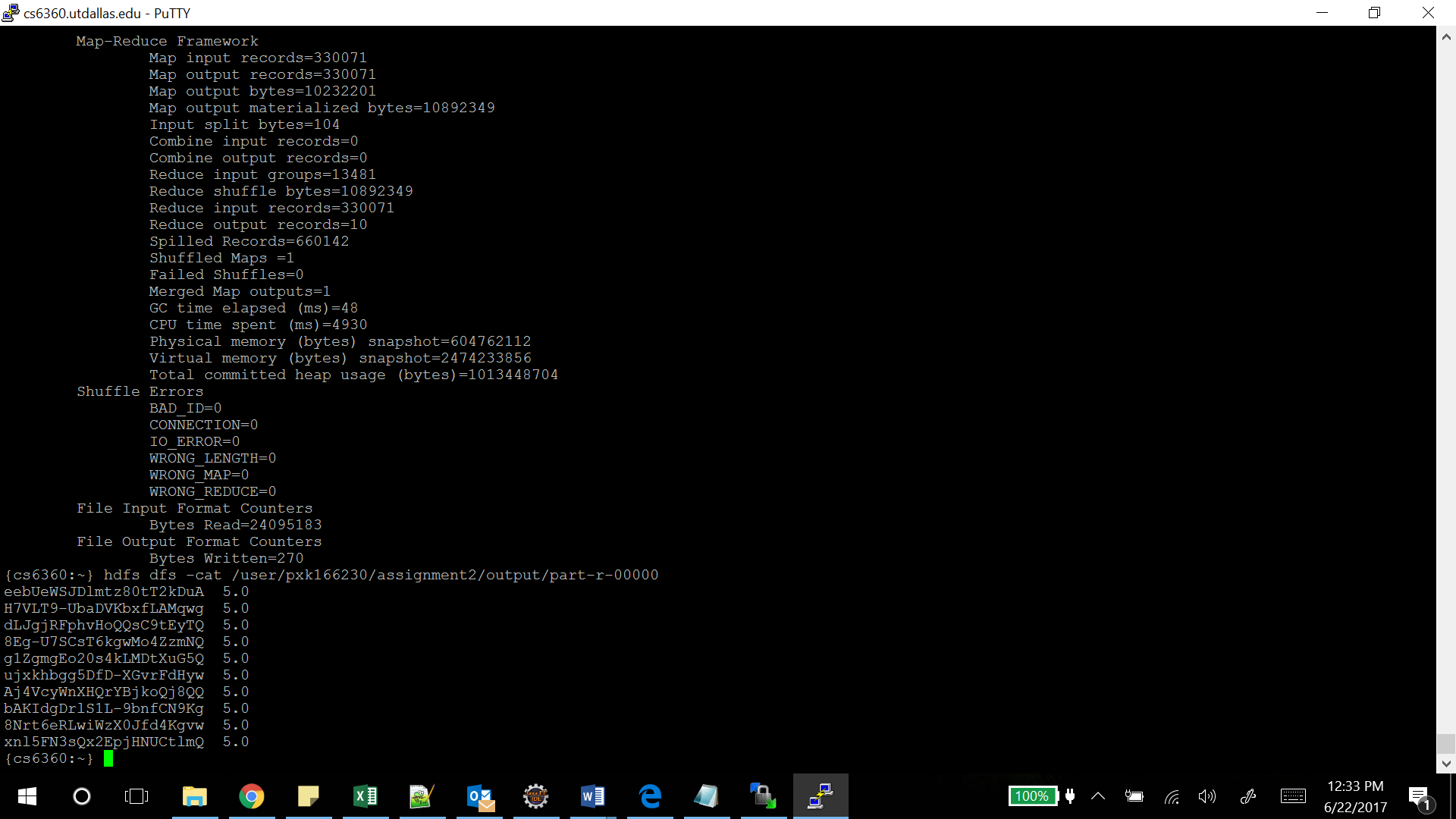
Q4:

*hadoop jar Assignment2.jar Q4 <input-path> <output-path>*

Example,

hadoop jar Assignment2.jar Q4 /yelp/review/review.csv hdfs://cshadoop1/user/pxk166230/assignment2/Q4/output

Screenshot of the output:



Q5:

*hadoop jar Assignment2.jar Q4 <input-path> <output-path>*

Example,

hadoop jar Assignment2.jar Q5 /yelp/review/review.csv hdfs://cshadoop1/user/pxk166230/assignment2/Q5/output

Screenshot of the output:

